13 Jan 2025 – Introduction to Statistics – Week 02

Estimators vs. estimate

point estimator?

Estimators	estimate
algebraic formula / function	numerical value
→ $\bar{x} = \frac{\sum x_i}{n}$	→ value of $\bar{x}$
→ does not change	→ changes with samples
	→ observed value of estimator

Unbiased estimator

→ $\text{Bias}(T) = E(T) - \theta$

e.g.: sample from class activity : $\theta = \mu$

$\overline{x_1} - \mu$, $\overline{x_2} - \mu$, ..., $\overline{x_{500}} - \mu$ } biases

$\text{avg}(\bar{x} - \mu) = E(\bar{x} - \mu) = E(T - \theta)$

In practice, θ is unknown, and only one sample is taken

→ Bias is a theoretical concept, not calculated in practical cases
only one sample dataset

Variance of an estimator

→ using mean is not sufficient to describe the population.

→ we also need to know about the spread

→ $V(T) = E(T - E(T))^2$

Mean square error

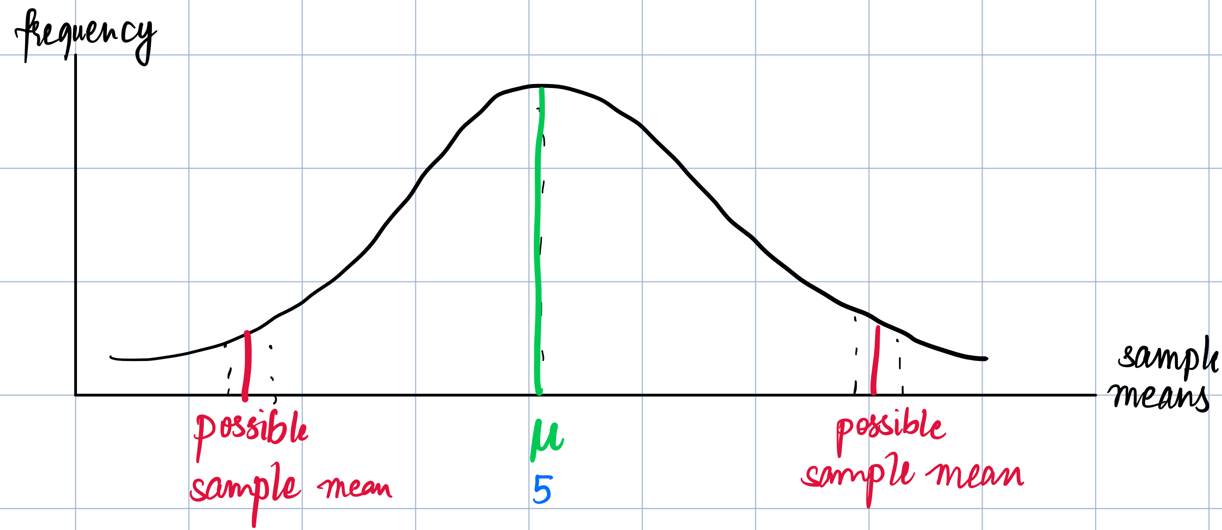
$MSE(T) = E(T - \theta)^2 = V(T) + Bias$

mean vs median
↓
better when outliers / extreme value

|||

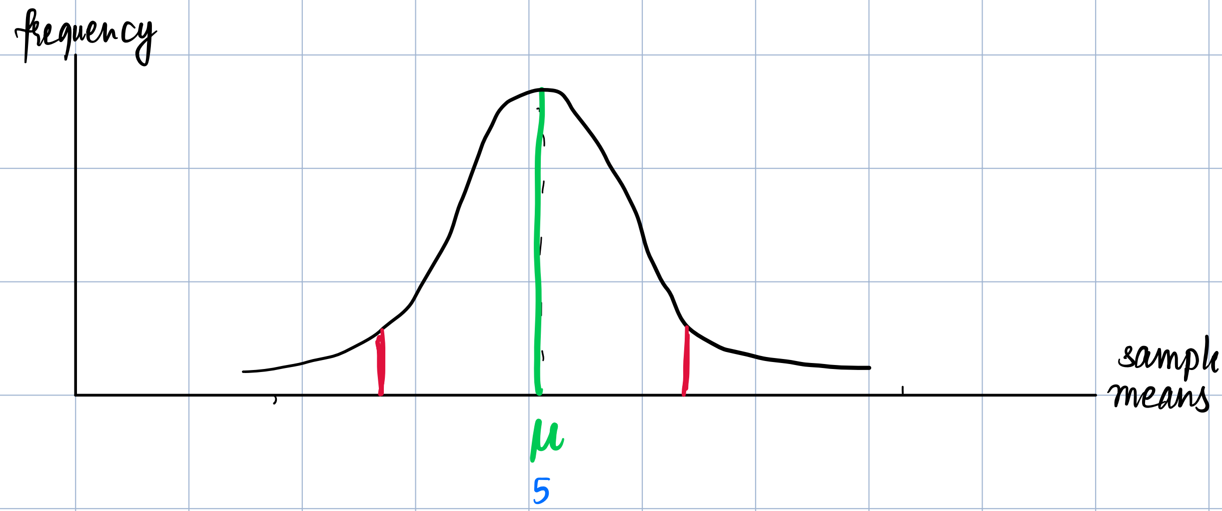
variance vs quartile deviation ??

5 ± 10



Preferable

5 ± 2



Best possible estimator : unbiased + no variance

(no variance: not possible)

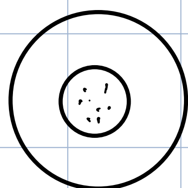
Minimum Variance Unbiased Estimator (MVUE)

$\bar{x} \rightarrow$ BLUE : Best Linear Unbiased Estimator

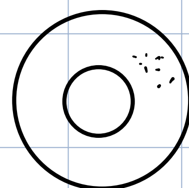
(Best → least possible variance + unbiased)

Accuracy and Precision

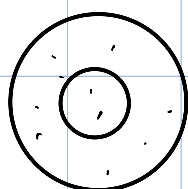
accurate and precise



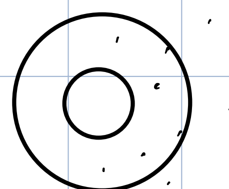
inaccurate but precise



accurate but imprecise



inaccurate and imprecise



e.g. traffic

route 1: 50 mins

route 2: 30 ± 40 mins

→ estimators that are more precise / (less variance) / predictable are preferred over estimators that are accurate but imprecise.

Class activity observations

1. bell curves
2. larger n lower variance, $n \uparrow$ $\sigma^2 \downarrow$
3. histograms of estimator
4. $n \uparrow$ precise
5. $E(\bar{x}) = \mu$

15 Jan 2025